

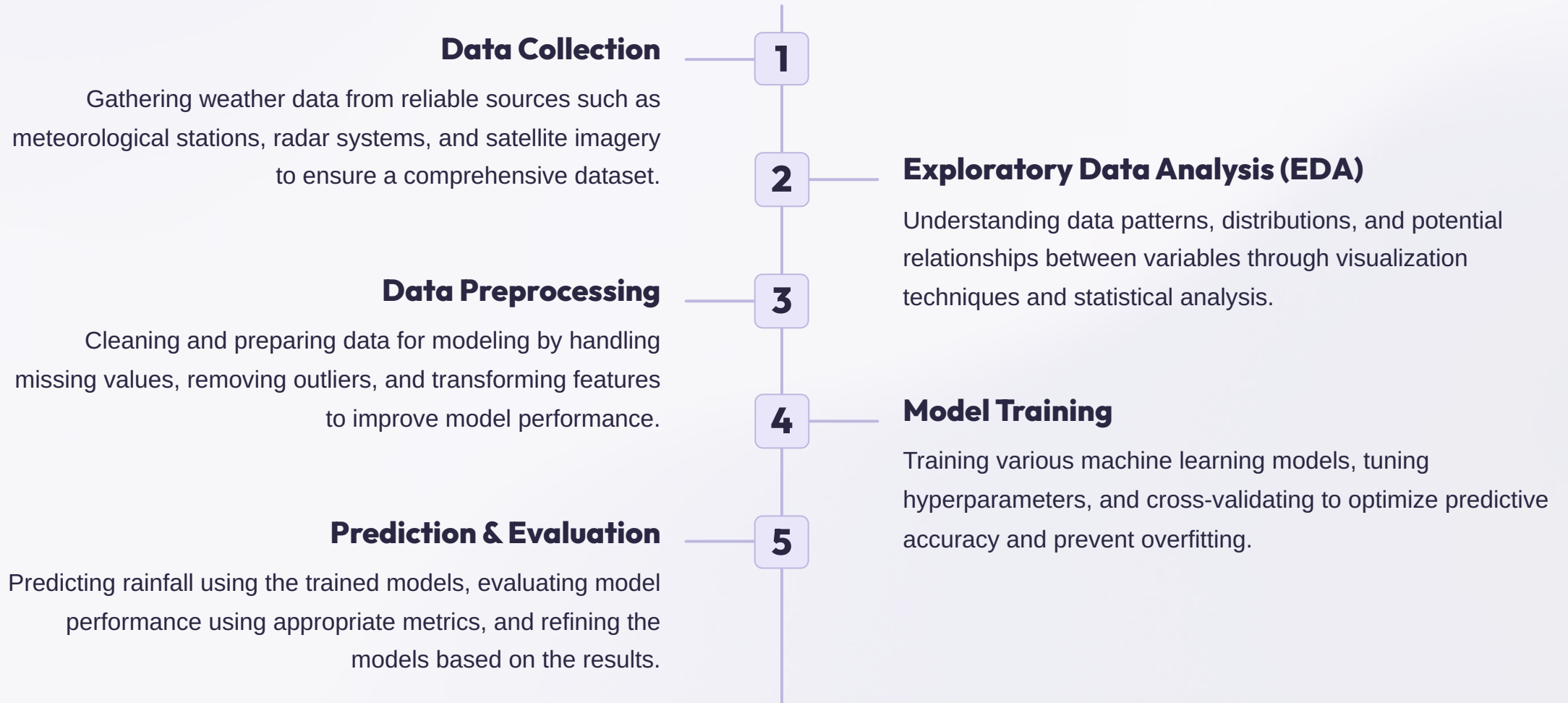
Predicting Rainfall with Machine Learning

This presentation outlines a machine-learning approach to weather prediction. The core objective of this project is to accurately forecast rainfall by leveraging various classification models. We'll cover each critical step of the workflow, from initial data collection and preprocessing to exploratory data analysis and, ultimately, comprehensive model evaluation. Our aim is to build a robust and reliable system that provides timely and accurate rainfall predictions. Let's begin by exploring the data collection process!

 by Sankari C



Project Workflow



Data Collection & Preprocessing

Loading the Dataset

Load data using pandas.

Cleaning the data

Remove whitespace

Handling Missing Values

Filling missing data using the mode and median.

Converting categorical columns

Transforming into numerical form for model compatibility.

Exploratory Data Analysis (EDA)

Statistical Measures

Calculate descriptive statistics to understand data distribution.

Distribution Analysis

Histograms to visualize feature distributions.

Correlation Matrix

Heatmap to identify feature correlations.

Addressing Data Imbalance



Targeted Data Analysis

Recognizing the imbalanced nature of rainfall data.



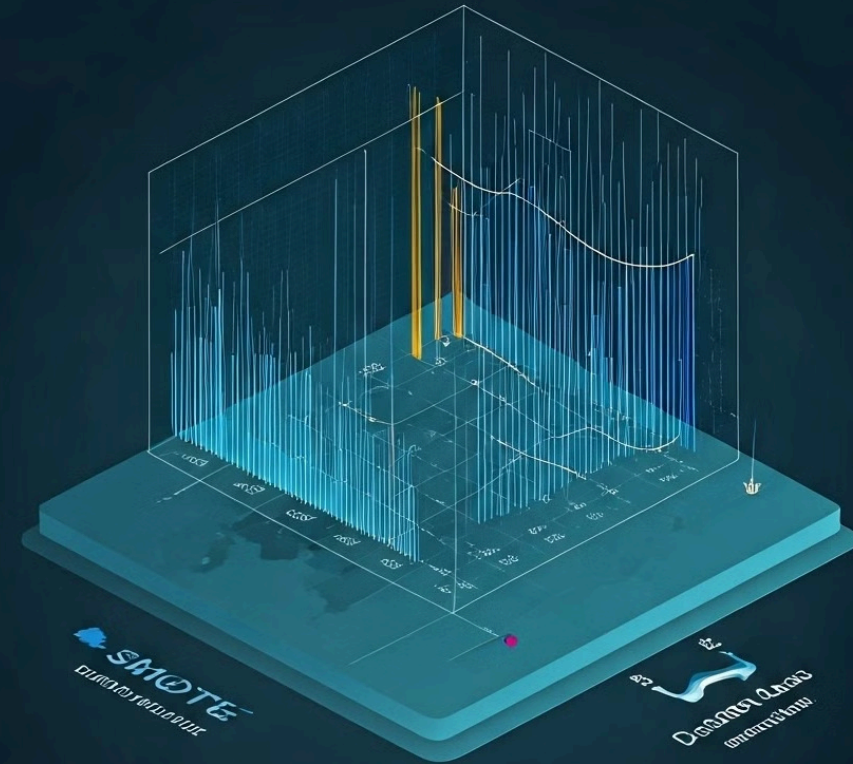
Balanced Data

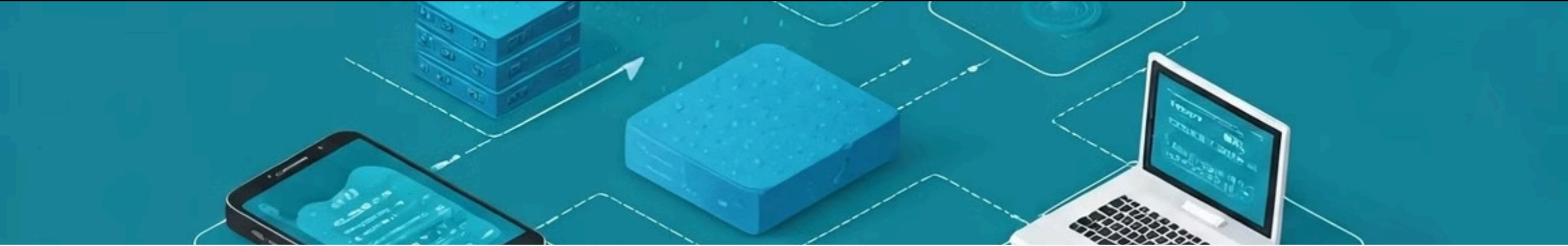
Ensure even data representation.



SMOTE Oversampling

Utilizing SMOTE to balance the dataset.





Splitting and Shuffling Data

1

Feature and Target Separation

Splitting the dataset into features and target variables.

2

Data Shuffling

Improve model performance by shuffling data randomly.

3

Training and Testing Split

Dividing the data for training and evaluation.

Logistic Regression



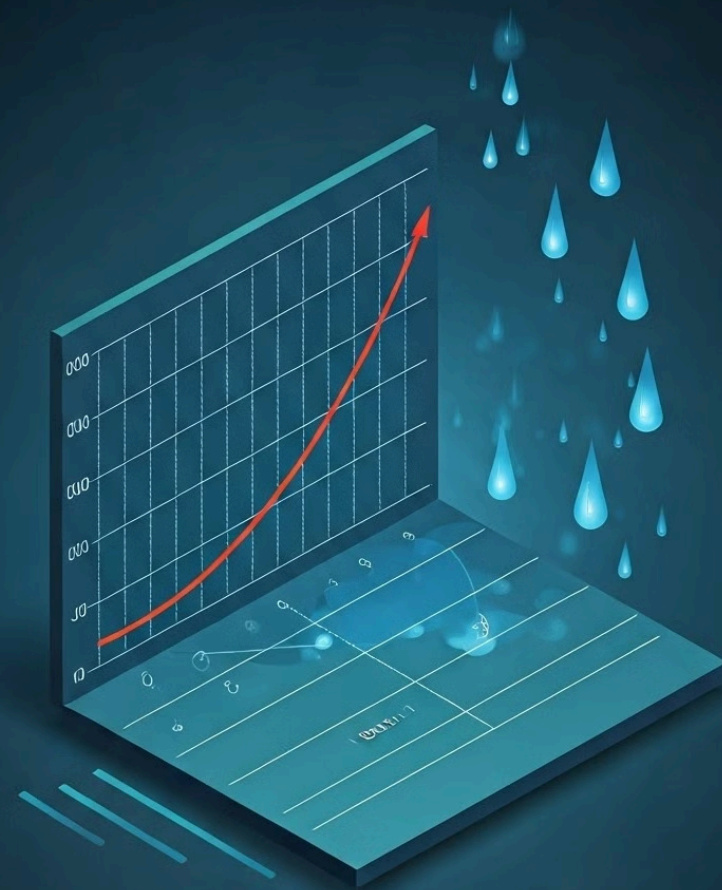
Hyperparameter Tuning



Accuracy

0.79 accuracy score with
Logistic Regression.

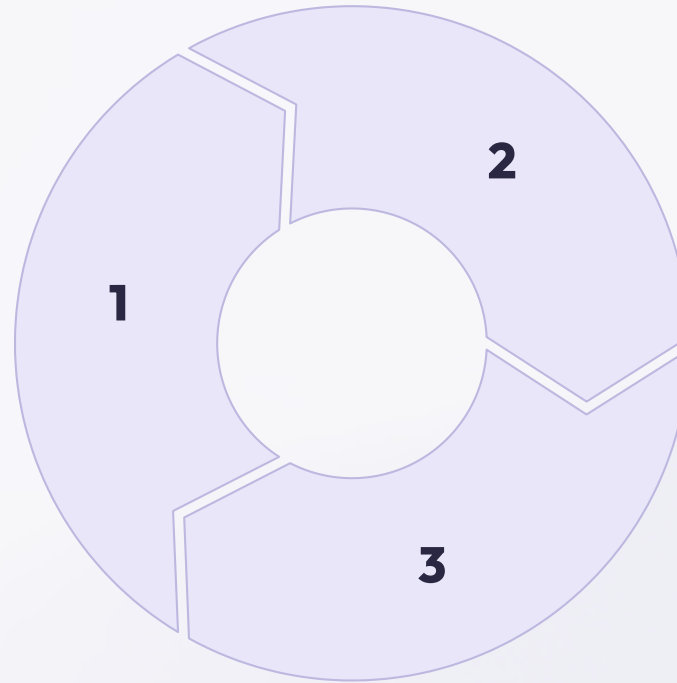
Logistic Regression employed with hyperparameter tuning. Achieved an accuracy score of **0.79**. This model provides a baseline for comparison.



XGBoost Classifier

Implementation

Implementing the XGBoost Classifier involves setting up the model with initial parameters and preparing it for training. Key steps include defining the objective function and evaluation metrics.



Tuning

Hyperparameter tuning is crucial for optimizing XGBoost. Techniques such as grid search or randomized search are used to find the best combination of parameters to enhance model performance.

Evaluation

Evaluating the XGBoost model involves using appropriate metrics like accuracy, precision, recall, and F1-score. These metrics help to quantify the model's performance on the test dataset and identify areas for improvement.

XGBoost Classifier model improved accuracy. Achieved a score of **0.81**, a notable improvement. XGBoost demonstrates better performance.

SVC Model Performance

1

SVC Model

Implementation and Configuration

2

Lowest Accuracy Score

Achieving only 0.49

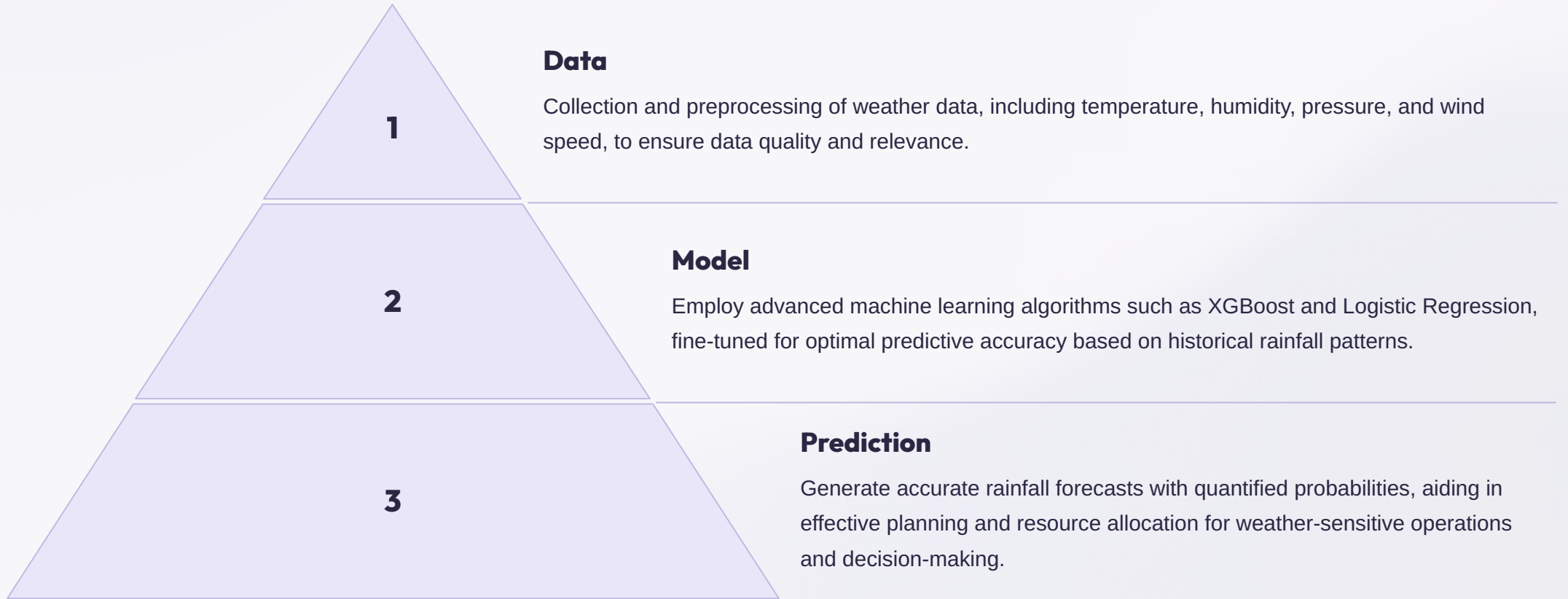
3

Unsuitable for Prediction

Limitations in capturing rainfall patterns

The SVC (Support Vector Classifier) model yielded the lowest accuracy score of 0.49, making it unsuitable for reliable rainfall prediction. This poor performance highlights the limitations of SVC in capturing the complex, non-linear relationships inherent in rainfall data. Further investigation into model parameters or alternative algorithms is necessary for improved predictive accuracy.

Prediction System



We have developed a comprehensive prediction system for rainfall, utilizing advanced machine learning techniques. By inputting relevant weather data, the system generates accurate rainfall predictions, enabling proactive planning and decision-making. This provides valuable, actionable weather insights for various applications.